

ESTIMATES AND PROJECTIONS AREA METHODOLOGY

COUNTY POPULATION ESTIMATES BY AGE, SEX, RACE, AND HISPANIC ORIGIN FOR JULY 1, 2002

BACKGROUND

The U.S. Census Bureau produces estimates of the resident population by age, sex, race and Hispanic origin for 3,141 counties in the United States on an annual basis. The following documentation outlines the methodology that was used in the production of the July 1, 2002 resident population estimates by age, sex, race, and Hispanic origin for all counties in the United States.

OVERVIEW

The Census Bureau develops total population estimates for counties with a cohort-component method. This method essentially follows each birth cohort according to its exposure to mortality, fertility, and migration. The cohort-component method is based on the traditional demographic accounting system and is described in more detail below. A major assumption underlying this approach is that the components of population change can be closely approximated by administrative data in a demographic change model. In order to apply the model, Census Bureau demographers estimate each component of population change separately. For the population residing in households the components of population change are births, deaths, and net migration, including net international migration. For the non-household population, change is represented by the net change in the population living in group-quarters facilities.

Each component in our model is represented with administrative data that are symptomatic of some aspect of population change. For example, birth certificates are symptomatic of additions to the population resulting from births, so we use these data to estimate the birth component for a county. Other components are derived from death certificates, Internal Revenue Service (IRS) data, Medicare enrollment records, Armed Forces data, group quarters population data. Data derived from the American Community Survey (ACS), Social Security files, Census 2000, and other internal Census Bureau sources are used to estimate some of the demographic details (age, sex, race, and Hispanic origin) for counties.

METHOD

The cohort-component method is based on the traditional demographic accounting system. Starting with a base population, deaths are subtracted from the population and births are added to the population, forming new cohorts. Estimates of net international migration and net internal migration are added to or subtracted from the population. The components of change are measured separately by age, sex, race, and Hispanic origin for each county and added to the base population as follows:

$$P_1 = P_0 + B - D + NDM + NIM$$

Where:

P_1 = population at the end of the period
 P_0 = population at the beginning of the period
 B = births during the period
 D = deaths during the period
 NDM = net internal migration during the period

NIM = net international migration during the period

As described above, administrative records, such as birth and death records, are used to estimate the components of population change. Net migration is calculated using several components including: net internal migration, net foreign-born international migration, net movement to/from Puerto Rico, net movement of federal and civilian citizens, the change in group-quarters population, and native emigration from the United States. Because the group quarters (GQ) population experiences somewhat different demographic processes, the GQ population is removed from the resident base population and estimated separately.

In the process of developing the July 1, 2002 estimates, revised estimates of the July 1, 2001 and July 1, 2000 county population with demographic detail were produced. The revised estimates for 2001 and 2000 incorporate actual data for the demographic components that were not previously available and includes updates or corrections to the data previously used. In cases where we do not have data for all counties for the current estimate year (2002), we estimate the components of population change based on one or more simplifying assumptions. When we develop the current population estimates, we use the same variant of the component model with these simplifying assumptions. In the creation of subsequent population estimates, we will replace these population estimates with "revised" population estimates based on actual data not yet received or corrected data.

One of the guiding principles in the Census Bureau's subnational methodology is that all of our population estimates are consistent. This means that the sum of the county estimates must be equal to the independently produced state and national population estimates. This consistency is required for all demographic characteristics produced. While this consistency is essential in the production and interpretation of the population estimates, it does add an additional layer of complexity to their development.

The methodology used to produce the July 1, 2002 estimates is described next.

STEP 1: SPECIFICATION OF THE BASE POPULATIONS

The first step was to subtract the GQ population from the Census 2000 resident population to develop a base population that consists of two pieces (the household population and the population residing in group quarters). Both pieces of the base population contain full demographic detail (age, sex, race, and Hispanic origin) for each county and the District of Columbia.

1A. Base Household Population

1A.i. The Census 2000 household population (obtained by subtracting the GQ population from the resident population) is the starting point for the July 1, 2002 county population estimates. The inclusion of demographic detail in the development of the county population estimates adds an additional layer of complexity to the estimation method. Census 2000 was the first census to allow multiple responses to the race question, but the administrative data sources used to estimate the components of change (births, deaths, and migration) were not available for all 31 races.¹ Therefore,

¹ The 31 race combinations include single responses for White, Black or African American, American Indian and Alaskan Native, Asian, and Native Hawaiian and other Pacific Islander; and all combinations of two or more of the five race groups. From this point on, the response Black or African American will be referred to as Black, and the response Native Hawaiian and other Pacific Islander will be referred to as Pacific Islander.

the Census 2000 base household population was converted from the 31 race combinations to the four race groups consistent with the 1990 Census. Then the July 1, 2002 population estimates were produced for the four race categories consistent with the 1990 Census. Finally, the July 1, 2002 estimates were converted to 31 races to be consistent with Census 2000 (see Step 6).

The conversion of Census 2000 race categories to 1990 Census race categories was based on proportional allocation. This assumes that multiple race responses in Census 2000 would be evenly distributed between the corresponding single race responses allowed in the 1990 Census. For example, the Census 2000 population in the three race categories of "White Alone," "Black Alone," and "White and Black" were converted into two of the 1990 Census race categories, "White" and "Black." The entire White Alone population enumerated in Census 2000 was assigned to White category and the entire Black Alone population was assigned to the Black category. Based on the assumption of straight proportion allocation, half of the Census 2000 "White and Black" population were assigned to the White race category and half of the population were assigned to the Black race category. These assignments were done at the county level, by sex and Hispanic origin.

The assumption of proportional allocation is the best available assumption at this time, though future estimates may not require this conversion or may be based on different distributions to the single race categories. See Step 6 for the conversion of the July 1, 2002 estimates for four races back to the 31 race categories.

1A.ii. Because the Census 2000 reference date is April 1, 2000 and the estimate periods are July 1 to June 30, a base population was calculated for July 1, 2000 using the July 1, 2000 state estimates with full demographic detail (by age, sex, race, and Hispanic origin) and July 1, 2000 county estimates for the age groups 0 to 64 and 65 and over. A ratio method was used to calculate July 1, 2000 county population estimates by age, sex, race, and Hispanic origin. This method applied the age, sex, race, and Hispanic origin distribution for counties from Census 2000 to the July 1, 2000 county population estimates for the two age groups to develop initial July 1, 2000 estimates with demographic detail. The state estimates by age, sex, race, and Hispanic origin were applied as controls to the initial estimates to generate county characteristics estimates that sum to equal the state distribution of demographic characteristics and the total population for counties for ages 0 to 64 and 65 and over.

The July 1, 2000 estimates then serve as the base population for the July 1, 2001 and July 1, 2002 estimates produced using the cohort-component method. The July 1, 2000 estimates could not be calculated using the cohort-component method because the administrative records used in the cohort-component method are available for calendar years and not the three-month time period from April 1, 2000 to July 1, 2000.

1B. Base Group-quarters Population

The group-quarters (GQ) population component is primarily a combination of military personnel living in barracks, college students living in dormitories, and persons residing in institutions. Inmates of correctional facilities, persons in health care facilities, persons in Job Corps Centers, and persons residing in nursing homes are also included in this category.

1B.i. The Census 2000 group quarters population is the starting point for the July 1, 2002 county GQ population estimates. First, the Census 2000 GQ data for the 31 race groups were converted to the four race groups consistent with the 1990 Census as described above.

1B.ii. The base group-quarters population for the July 1, 2002 estimates was the revised group-quarters population from July 1, 2001. The July 1, 2001 GQ estimates were calculated starting with the Census 2000 GQ population by age, sex, race, Hispanic origin, and seven group quarters types for each county. States provide updated information on the total GQ population by GQ type to the Census Bureau each year. The Census 2000 age, sex, race, and Hispanic origin distributions of the GQ population by county and type were applied to the July 1, 2001 and July 1, 2002 GQ populations reported by the Federal-State Cooperative Program for Population Estimates (FSCPE) to produce GQ population estimates for counties with demographic detail.

STEP 2: SPECIFICATION OF BIRTHS AND DEATHS (VITAL STATISTICS) COMPONENTS

2A. The birth and death components are calculated from three sources of data. Files containing all registered births and deaths that occurred to U.S. residents during the estimate period are obtained from the National Center for Health Statistics (NCHS). The birth files contain the total numbers of births in a calendar year by state and county of mothers' residence, sex, race, and Hispanic origin. The NCHS death files contain the total numbers of deaths by sex, race, Hispanic origin, age at death, and state and county of residence at death. The FSCPE also report annual numbers of registered births and deaths by sex, race, Hispanic origin, age at death, and state and county of residence at death or county of mothers residence at birth. A reconciliation process occurs between the NCHS and FSCPE vital statistics. In general, we believe that the total demographic characteristics distribution of data from the NCHS file is more accurate due to its national coverage, while the geographic distribution of data from FSCPE files are more accurate due to the FSCPEs more specific local knowledge.

It is assumed that the vital statistics files represent complete counts of births and deaths for the resident population. No adjustments are made for undercoverage or differential coverage by county, age, race, or Hispanic origin.

2B. After the NCHS and FSCPE figures are reconciled, they are controlled to the national estimates of the numbers of births and deaths by sex, race, Hispanic origin, and age at death developed as part of the national population estimates for the same time period. 2C. Finally, the births are added to the base population for each year (July 1, 2000 and July 1, 2001) and the deaths are subtracted from the base population by county and demographic characteristics. As the vital statistics are provided for calendar years, assumptions are made to apply the births and deaths to the July 1 reference dates.

STEP 3: SPECIFICATION OF NET INTERNATIONAL MIGRATION

We estimate the net international migration to/from the United States from several sources. Our estimate includes the net foreign-born international migration, net movement to/from Puerto Rico, net federal and civilian citizen movement, and native emigration. With the exception of military station strength data, the majority of these data are developed first at the national level (e.g., international migration components, totals by age, sex, race, Hispanic origin). These national and demographic characteristics totals are distributed to states and counties using Census 2000 proportions.

National-level July 1, 2000, July 1, 2001 and July 1, 2002 estimates of the net foreign-born international migration for each year are distributed to counties based on the distribution of the foreign-born population who entered the U.S. during the 5 years prior to April 1, 2000 by country of birth as measured by Census 2000.

National-level July 1, 2000, July 1, 2001 and July 1, 2002 estimates of the total net movement of the population to or from Puerto Rico for each year are distributed to counties based on the distribution of the Puerto Rican population in Census 2000.

The number of Armed Forces personnel stationed at military bases is supplied by each branch of the Armed Forces. County-level July 1, 2000, July 1, 2001 and July 1, 2002 estimates of the total federal and civilian citizen movement of the population are derived by applying the Census 2000 distribution of Armed Forces personnel by age, sex, race, and Hispanic origin to the county-level station strength data.

National-level July 1, 2000, July 1, 2001 and July 1, 2002 estimates of the total number of foreign-born emigrants from the United States for each year are distributed to counties based on the distribution of the foreign-born population in Census 2000 by country of birth.

STEP 4: SPECIFICATION OF NET INTERNAL MIGRATION

Step 4A. Match of Tax Returns to create counts of exemptions (filers and dependents) who migrate by demographic characteristics

4A.i. For the July 1, 2002 estimates the component of internal migration was developed using data from two administrative record sources: annual extracts of tax returns provided by the Internal Revenue Service (IRS) linked by Social Security Number across successive years; and the Census Numident file, derived from the Social Security Administration 100 percent file (SSA). In order to ensure confidentiality and privacy, these data sets are matched by SSN/PIK (Protected Identification Key). These data are referred to jointly as IRS-SSA data. The IRS 1040 tax return records were matched to the SSA data to identify the age, sex, race, and Hispanic origin of the tax filers. A number of assumptions were made to assign demographic characteristics to spouses and dependents. Exemptions claimed for children were assigned to the under 20 age group and exemptions claimed for parents were assigned to age 65 and over. Sex was assigned randomly for exemptions. Spouses were assigned the same age and the opposite sex as filers. All spouses and exemptions were assigned the same race and Hispanic origin as filers.

4A.ii. After the demographic characteristics were added to the IRS tax return records, two years of records were matched by SSN/PIK to determine the migration status. Filers (and their dependents) with a change in the state of residence between the two periods were identified as "Inter-State" migrants. Filers (and their dependents) with a change in the county of residence (but not their state of residence) between two periods were identified as "Intra-State" migrants. Otherwise, if there was no change in the state or the county of residence, the filers (and dependents) were identified as non-migrants.

Because of the potentially large number of origin-characteristic combinations, a few simplifying assumptions were required in the production of the July 1, 2002 estimates. Inter-state and intra-state migrants were calculated separately as documented below. Further, it was necessary in some cases to combine individual origin-characteristic categories (which will be referred to as cells) to improve the robustness of the data. If a given cell had less than 30 exemptions, then it was combined with adjacent age cells within the same origin-ethnicity-race-sex group until the combined category contained at least 30 exemptions. If it was not possible to create a combined category containing at least 30 exemptions within an origin-ethnicity-race-sex group, then cells were combined for both sexes. When single years of age were combined to compute a

migration probability, each of the ages was assigned the probability for the aggregated age group.

Step 4B. Calculate Inter-State Out-Migration Rates and Inter-State Out-Migrant Population
The inter-state out-migration probability was calculated as the ratio of the total number of inter-state migrant exemptions from that state divided by the total number of exemptions (migrants and non-migrants) in the same state. These rates were calculated by age, sex, race, and Hispanic origin. The estimated number of inter-state out-migrants was calculated by multiplying the inter-state out-migration rate for each set of demographic characteristics for each state by the base household population for each state (see Step 1). (Are GQ returns ever treated separately?)

Step 4C. Calculate In-Migration Proportion and In-Migrant Population
From the matched records the destinations of migrants by demographic characteristics can be determined. The numbers of inter-state out-migrants calculated from Step 4B were distributed as in-migrants to states by applying proportions of total in-migrant exemptions who moved to each state. These proportions were calculated by age, sex, race, and Hispanic origin. The numbers of out-migrants by state and characteristics were subtracted from the base population for each time period and the numbers of in-migrants were added to the base population.

Step 4D. Calculate Intra-State Out/In-Migration Proportions and Intra-State Out/In Migrant Population
While it is theoretically possible to handle intra-state migration in the same manner as inter-state migration, the combinations of 3,141 counties and the demographic characteristics would require millions of cells, many with too few observations to produce reliable estimates. Thus, for practical purposes, intra-state migration was handled differently. Both intra-state out-migration and intra-state in-migration were calculated as proportions of the total intra-state migrant exemptions by demographic characteristics for the state. Thus, the first step was to calculate the county's share of the state's total intra-state out-migration exemptions by demographic characteristics. Then the county's share of the state's total intra-state in-migration exemptions by demographic characteristics was calculated. These proportions were then applied to the county base population to calculate the population of intra-state out-migrants and intra-state in-migrants. Finally, the estimated numbers of in- and out-migrants were added to and subtracted from the county base populations.

STEP 5: PROCESSING OF GROUP QUARTERS POPULATION

The GQ population change was estimated separately from the demographic accounting procedure for the household population estimates described above. This was done primarily because of the uniqueness of this subpopulation and the special difficulties of estimating the GQ population. The Census 2000 GQ data has full demographic detail (age, sex, race, and Hispanic origin) and information on type of GQ residence classified into seven types: (1) Correctional Facilities; (2) Juvenile Institutions; (3) Nursing Homes; (4) Other Institutional Facilities; (5) College Dorms; (6) Military Barracks; and (7) Other Non-Institutional GQ. The GQ population is updated for each estimate year using information on the total GQ population by GQ type collected by state agencies through the FSCPE. As with the July 1, 2001 base population (see Step 1B), the Census 2000 age, sex, race, and Hispanic origin distributions of the GQ population by county and type were applied to the July 1, 2002 total GQ populations for counties to produce GQ population estimates for counties with demographic detail.

STEP 6: GENERATE RESIDENT POPULATION ESTIMATES BY DEMOGRAPHIC CHARACTERISTIC

6A. Prior to combining the July 1, 2000 (revised), July 1, 2001 (revised), and July 1, 2002 group quarters and household population estimates, each set of estimates were converted from the four race groups consistent with the 1990 Census to the 31 race groups consistent with Census 2000 by applying conversion factors. Continuing the example from Step 1A.i., the estimated "White" population was apportioned to the "White Alone" and "White and Black" populations. To produce the July 1, 2002 estimate for the White Alone population, the estimated July 1, 2002 White population was multiplied by the ratio of the White Alone population from Census 2000 to the sum of the White Alone and half of the "White and Black" population from Census 2000. To produce part of the July 1, 2002 estimates for the "White and Black" population, the estimated July 1, 2002 White population was multiplied by the ratio of the "White and Black" population from Census 2000 to the sum of the White Alone and half of the "White and Black" population from Census 2000. The remaining part of the July 1, 2002 "White and Black" population was obtained by applying comparable ratios to the July 1, 2002 estimates Black population.

6B. GQ and household estimates of the population were summed by county, age, sex, race, and Hispanic origin within counties. The preliminary estimates were compared with independently calculated county total population estimates, and state estimates by age, sex, race and Hispanic origin. The final resident population estimates were adjusted to equal the independent totals by multiplying each preliminary estimate by the ratio of the independent totals to the sum of the relevant preliminary estimates. These adjusted estimates were rounded to whole numbers for each combination of demographic characteristics within counties and compared with the independent totals.

In our quality control checks, we noted 146 observations where the summed county characteristics did not equal exactly the separately produced national characteristics estimates, while the sums of 31,846 other county characteristics combinations equaled the independently produced totals exactly. This was a product of controlled rounding. The largest discrepancies were: -566 and 689 for July 1, 2000 estimates; -504 and 650 for July 1, 2001 estimates; and -438 and 584 for July 1, 2002 estimates. For each year of estimates, the mean of the differences was zero. Further examination showed that these discrepancies occurred for age 84 years and the age group 85 and over, and were largest for the White Alone and the "White and Black" race groups.